

# **IMPLEMENTASI DATA MINING UNTUK MENGETAHUI PERILAKU SESEORANG DALAM MEMBELI ALAT PANAHAH**



## **PUBLIKASI ILMIAH**

**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada Program Studi  
Informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**ARMANDO PRIYA TARUNA**

**L200100153**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**2017**

**HALAMAN PERSETUJUAN**

**IMPLEMENTASI DATA MINING UNTUK MENGETAHUI  
PERILAKU SESEORANG DALAM MEMBELI ALAT  
PANAHAH**

**PUBLIKASI ILMIAH**

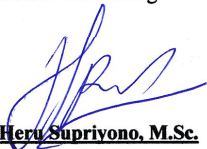
oleh:

**ARMANDO PRIYA TARUNA**

**L200100153**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



**Dr. Heru Supriyono, M.Sc.**

**NIK 970**

**HALAMAN PENGESAHAN**  
**IMPLEMENTASI DATA MINING UNTUK MENGETAHUI**  
**PERILAKU SESEORANG DALAM MEMBELI ALAT**  
**PANAHAN**

OLEH

**ARMANDO PRIYA TARUNA**

**L200100153**

Telah dipertahankan di depan Dewan Penguji  
Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Sabtu, 2 Desember 2017  
dan dinyatakan telah memenuhi syarat

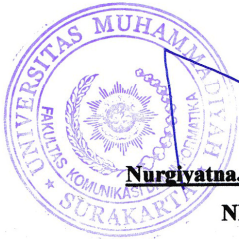
Dewan Penguji:

1. **Dr. Heru Supriyono, M.Sc.**  
(Ketua Dewan Penguji)
2. **Nurgiyatna, S.T, M.Sc, Ph.D**  
(Anggota I Dewan Penguji)
3. **Yogiek Indra Kurniawan, S.T, M.T.**  
(Anggota II Dewan Penguji)

(.....)  
(.....)  
(.....)

Dekan,

Ketua Program Studi,



**Nurgiyatna, S.T, M.Sc, Ph.D.**

**NIK 881**



**Dr. Heru Supriyono, M.Sc.**

**NIK 970**

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam tugas pengganti skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

**Surakarta, 2 Desember 2017**

Penulis



**ARMANDO PRIYA TARUNA**

**L200100153**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

No surat .....396/A.3-11.7/INF-FKI/XII/2017

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : Armando Priya Taruna  
NIM : L200100153  
Judul : **IMPLEMENTASI DATA MINING UNTUK MENGETAHUI  
PERILAKU SESEORANG DALAM MEMBELI ALAT PANAHAHAN**  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 8 Desember 2017

Biro Skripsi Informatika

**Ihsan Cahyo Utomo, S.Kom., M.Kom.**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA**  
**FAKULTAS KOMUNIKASI DAN INFORMATIKA**  
**PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

turnitin

IMPLEMENTASI DATA MINING UNTUK MENGETAHUI PERILAKU SESEORANG DALAM MEMBELI ALAT PANAHAN

Armando Priya Taruna, Heru Supriyono

**Abstrak**

Olahraga panahan sekarang ini sudah semakin banyak diminati, ada yang ingin menjadi atlet ada juga yang hanya sekedar hobi, namun dengan harga alat yang mahal menjadi kendala bagi orang yang akan menekuni olahraga ini. Untuk membantu orang dalam memilih kondisi alat panahan yang akan dibeli, membantu penjual alat panahan dalam menentukan target konsumen, maka perlu dilakukan pengambilan keputusan yang tepat. Dengan *data mining* klasifikasi algoritma C4.5 diharapkan dapat mencari tahu bagaimana perilaku konsumen dalam membeli alat panahan. Data masukan diperoleh dari data orang yang mengikuti latihan panahan di GOR Diponegoro Kota Surab. Data yang diperoleh berupa data harga alat panahan, motivasi dalam mengikuti latihan panahan dan penghasilan. Hasil pengolahan data berupa pohon keputusan. Motivasi menjadi faktor utama dalam membeli alat panah baru, terutama motivasi menjadi seorang atlet, untuk motivasi hanya sebagai hobi, penghasilan dan harga menjadi faktor penentu untuk membeli alat panah baru atau tidak. Penerapan *data mining* algoritma C4.5 dapat membantu penjual alat panahan untuk mengetahui perilaku konsumen dalam membeli alat panahan terutama di Kota Surab.

Page: 1 of 13 Word Count: 2660

Match Overview

18%

|   |  |                 |    |
|---|--|-----------------|----|
| 1 | <a href="http://eprints.ums.ac.id">eprints.ums.ac.id</a>               | Internet Source | 3% |
| 2 | Submitted to Universita...   | Student Paper   | 3% |
| 3 | <a href="http://dokumen.tips">dokumen.tips</a>                         | Internet Source | 3% |
| 4 | <a href="http://eprints.ums.ac.id">eprints.ums.ac.id</a>               | Internet Source | 1% |
| 5 | <a href="http://es.scribd.com">es.scribd.com</a>                       | Internet Source | 1% |
| 6 | <a href="http://research.pps.dinus.ac.id">research.pps.dinus.ac.id</a> | Internet Source | 1% |
| 7 | <a href="http://media.neliti.com">media.neliti.com</a>                 | Internet Source | 1% |

## **IMPLEMENTASI *DATA MINING* UNTUK MENGETAHUI PERILAKU SESEORANG DALAM MEMBELI ALAT PANAHAN**

### **Abstrak**

Olahraga panahan sekarang ini sudah semakin banyak diminati, ada yang ingin menjadi atlet ada juga yang hanya sekedar hobi, namun dengan harga alat yang mahal menjadi kendala bagi orang yang akan menekuni olahraga ini. Untuk membantu orang dalam memilih kondisi alat panahan yang akan dibeli, membantu penjual alat panahan dalam menentukan target konsumen, maka perlu dilakukan pengambilan keputusan yang tepat. Dengan *data mining* klasifikasi algoritma C4.5 diharapkan dapat mencari tahu bagaimana perilaku konsumen dalam membeli alat panahan. Data masukan diperoleh dari data orang yang mengikuti latihan panahan di GOR Diponegoro Kota Sragen. Data yang diperoleh berupa data harga alat panahan, motivasi dalam mengikuti latihan panahan dan penghasilan. Hasil pengolahan data berupa pohon keputusan. Motivasi menjadi faktor utama dalam membeli alat panah baru, terutama motivasi menjadi seorang atlet, untuk motivasi hanya sebagai hobi, penghasilan dan harga menjadi faktor penentu untuk membeli alat panah baru atau tidak. Penerapan *data mining* algoritma C4.5 dapat membantu penjual alat panahan untuk mengetahui perilaku konsumen dalam membeli alat panahan terutama di Kota Sragen.

**Kata Kunci:** Algoritma C4.5, *Data Mining*, Klasifikasi, Minat Olahraga Panahan.

### ***Abstract***

*Archery sports are now increasingly in demand, there are who want to be an athlete there is also just a hobby, but with the price of expensive equipment becomes an obstacle for people who will pursue this sport. To assist people in choosing the condition of the archery to be purchased, help archery tool seller in determining consumer target, it is necessary to make the right decision. With data mining C4.5 algorithm classification is expected to find out how consumer behavior in buying the arrow tool. Data input obtained from the data of people who follow the archery training in GOR Diponegoro Sragen. Data obtained in the form of archery equipment price data, motivation in following the archery practice and income. The result of data processing is a decision tree. Motivation is a major factor in buying new arrow tools, especially the motivation of being an athlete, for motivation only as a hobby of income and price becomes the deciding factor to buy a new arrow tool or not. Implementation of data mining algorithm C4.5 can help archery tool seller to know consumer behavior in buying tool of archery especially in Sragen.*

**Keywords:** Algorithm C4.5, *Data Mining*, Classification., Archery Sports Interests.

## 1. PENDAHULUAN

Olahraga merupakan keharusan dan gaya hidup bagi sebagian orang, salah satu olahraga yang saat ini sedang populer dan diminati banyak orang adalah panahan, namun olahraga panahan sendiri merupakan olahraga yang bisa dibilang mahal, karena untuk membeli alat saja bisa menghabiskan biaya 1 juta rupiah sampai 10 juta rupiah tergantung kelas masing masing.

Mahal tidak berarti olahraga panahan sepi peminat, dari anak-anak sampai orang dewasa banyak yang tertarik dengan olahraga ini, hal ini bisa dilihat dari intensitas latihan dan banyaknya orang yang berlatih panahan, selain itu sekolah-sekolah mulai menjadikan olahraga panahan sebagai ekstrakurikuler dan kejuaraan-kejuaraan junior ataupun umum sudah sering diselenggarakan yang menjadi daya tarik lebih bagi olahraga ini,

Selain untuk menjadi atlet sebagian orang membeli alat panahan untuk menyalurkan hobi mereka, ada yang untuk melestarikan tradisi, ada juga yang untuk berburu dan untuk orang yang usianya sudah agak tua olahraga panahan bisa menjadi alternatif olahraga yang cukup baik, karena olahraga ini tidak membutuhkan fisik yang terlalu banyak dan bisa untuk melatih fokus mereka.

Banyak orang tua yang ingin memasukan anak mereka dalam olahraga panahan terutama di Kabupaten Sragen, namun mereka menganggap olahraga ini modalnya cukup besar, sebenarnya alat panahan bisa di beli dengan harga yang terjangkau, orang bisa membeli alat bekas atau membeli per bagian saja secara bertahap. Untuk membantu orang dalam memilih kondisi alat panahan yang akan dibeli, membantu penjual alat panahan dalam menentukan target konsumen, maka perlu dilakukan pengambilan keputusan yang tepat.

Dengan kemajuan teknologi seperti sekarang ini banyak cara yang dapat dilakukan untuk membantu dalam mengambil keputusan, salah satu caranya dengan melakukan *data mining*. *Data mining* dapat diartikan sebagai pengestrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *knowledge discovery*.



Salah satu metode yang digunakan adalah klasifikasi dengan algoritma C4.5. Dengan algoritma ini penulis mencoba mencari tahu bagaimana perilaku konsumen dalam membeli alat panahan.

## **2. METODE**

Untuk memproses data dan mengetahui perilaku seseorang dalam membeli alat panahan baru atau bekas menggunakan algoritma C4.5 dan untuk menghitung menggunakan *Microsoft Office Excel 2007*.

### **2.1 Pohon Keputusan**

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat terkenal. Pohon keputusan berguna untuk menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target, karena pohon keputusan memadukan antara eksplorasi data dan pemodelan.

Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan/internal (*branch/internal node*) dan simpul daun (*leaf node*).

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi aturan dan menyederhanakan aturan.

### **2.2 Algoritma C4.5**

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada. Sebelum menghitung *information gain*, kita perlu menghitung dulu entropi

Cara menghitung dilakukan dengan menggunakan konsep entropi. Untuk menghitung entropi dapat dilakukan dengan rumus (1) :

$$\text{Entropi}(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \quad ..(1)$$

Keterangan:

S = Himpunan kasus

P1 = Jumlah kasus pada partisi ke-1

P2 = Jumlah kasus pada partisi ke-2

Pn = Jumlah kasus pada partisi ke-n

Jika perbandingan dua kelas rasionya sama maka nilai entropinya 1. Jika satu set hanya terdiri dari satu kelas maka entropinya 0.

Setelah menghitung entropi kemudian menghitung *information gain*. *Information gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan *informasi gain* tertinggi dipilih sebagai *test atribut* dari suatu *node*. Untuk menghitung *information gain* dapat dilakukan dengan rumus (2) :

$$\text{Gain}(S, A) = \text{entropi}(S) - \sum_{c \in \text{nilai}(S)} \frac{Sc}{S} \text{entropi}(Sc) \quad ..(2)$$

Keterangan:

S = Himpunan kasus

A = Atribut

Sc = Jumlah kasus pada partisi ke-i

S = Jumlah kasus dalam S

### 3. HASIL DAN PEMBAHASAN

Pengumpulan data sendiri dilakukan dengan cara mendata atlet panahan Kabupaten Sragen dan melakukan wawancara kepada atlet ataupun orang-orang yang rutin ikut latihan panahan di GOR Diponegoro Kabupaten Sragen. Setelah data diperoleh maka akan ditentukan variabel-variabel yang akan menjadi keputusan dalam pembelian alat panahan. Penulis menentukan variabel yang

digunakan berupa data harga alat panahan, motivasi mengikuti latihan dan penghasilan. Untuk data hasil penelitian disajikan pada tabel 1.

Tabel 1. Data Sebelum Pemrosesan

| No | Harga     | Motivasi | Penghasilan | Beli  |
|----|-----------|----------|-------------|-------|
| 1  | 5 - 10 jt | Altet    | >5jt        | Baru  |
| 2  | <1 jt     | Atlet    | <1jt        | Baru  |
| 3  | 5 - 10 jt | Atlet    | 1 - 5jt     | Baru  |
| 4  | 5 - 10 jt | Hobi     | <1jt        | Bekas |
| 5  | 5 - 10 jt | Atlet    | 1 - 5jt     | Baru  |
| 6  | 5 - 10 jt | Atlet    | 1 - 5jt     | Baru  |
| 7  | <1 jt     | Atlet    | <1jt        | Baru  |
| 8  | <1 jt     | Hobi     | <1jt        | Bekas |
| 9  | 5 - 10 jt | Hobi     | <1jt        | Bekas |
| 10 | <1 jt     | Hobi     | 1 - 5jt     | Baru  |
| 11 | <1 jt     | Atlet    | 1 - 5jt     | Baru  |
| 12 | 5 - 10 jt | Atlet    | <1jt        | Baru  |
| 13 | <1 jt     | Atlet    | 1 - 5jt     | Baru  |
| 14 | 5 - 10 jt | Hobi     | <1jt        | Bekas |
| 15 | <1 jt     | Atlet    | <1jt        | Baru  |
| 16 | <1 jt     | Hobi     | 1 - 5jt     | Baru  |
| 17 | <1 jt     | Hobi     | <1jt        | Bekas |
| 18 | <1 jt     | Hobi     | <1jt        | Bekas |
| 19 | 5 - 10 jt | Atlet    | <1jt        | Baru  |
| 20 | 5 - 10 jt | Atlet    | >5jt        | Baru  |
| 21 | <1 jt     | Hobi     | 1 - 5 jt    | Baru  |
| 22 | 5 - 10 jt | Hobi     | 1 - 5 jt    | Bekas |
| 23 | <1 jt     | Hobi     | 1 - 5 jt    | Baru  |
| 24 | <1 jt     | Hobi     | <1jt        | Bekas |
| 25 | 5 - 10 jt | Hobi     | >5jt        | Baru  |

### 3.1 Menghitung Entropi dan *Information Gain*

#### 3.1.1 Penghitungan Iterasi 1

Untuk menghitung entropi pada atribut target, terlebih dahulu dilakukan pemisahan kelas yang mungkin yaitu beli baru dan beli bekas, untuk lebih jelasnya dapat dilihat pada tabel 2.

Tabel 2. Seleksi Atribut *Independent* Beli

| Beli | Baru | Bekas |
|------|------|-------|
|      | 17   | 8     |

$$\begin{aligned}
 Entropy(\text{Total}) &= - \left( \frac{17}{25} * \log_2 \left( \frac{17}{25} \right) \right) - \left( \frac{8}{25} * \log_2 \left( \frac{8}{25} \right) \right) \\
 &= 0,90438146
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut harga terebih dahulu dilakukan seleksi atribut *independent* pada atribut harga, untuk lebih jelasnya dapat dilihat pada tabel 3.

Tabel 3. Seleksi Atribut *Independent* Harga

| Harga  | Baru | Bekas |
|--------|------|-------|
| <1jt   | 9    | 4     |
| 5-10jt | 8    | 4     |

$$\begin{aligned}
 Entropy(\text{Harga}, <1\text{jt}) &= - \left( \frac{9}{13} * \log_2 \left( \frac{9}{13} \right) \right) - \left( \frac{4}{13} * \log_2 \left( \frac{4}{13} \right) \right) \\
 &= 0,89049164
 \end{aligned}$$

$$\begin{aligned}
 Entropy(\text{Harga}, 5-10\text{jt}) &= - \left( \frac{8}{12} * \log_2 \left( \frac{8}{12} \right) \right) - \left( \frac{4}{12} * \log_2 \left( \frac{4}{12} \right) \right) \\
 &= 0,91829583
 \end{aligned}$$

$$\begin{aligned}
 Information\ gain(\text{Total}, \text{Harga}) &= 0,90438146 - \left( \left( \frac{13}{25} * 0,89049164 \right) + \left( \frac{12}{25} * 0,91829583 \right) \right) \\
 &= 0,0005438
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut motivasi terebih dahulu dilakukan seleksi atribut *independent* pada atribut motivasi, untuk lebih jelasnya dapat dilihat pada tabel 4.

Tabel 4. Seleksi *Independent* Atribut Motivasi

| Motivasi | Baru | Bekas |
|----------|------|-------|
| Atlet    | 12   | 0     |
| Hobi     | 5    | 8     |

$$Entropy(\text{Motivasi}, \text{Atlet}) = 0, \text{ karena hanya terdiri dari satu kelas}$$

$$\begin{aligned}
 Entropy(\text{Motivasi}, \text{Hobi}) &= - \left( \frac{5}{13} * \log_2 \left( \frac{5}{13} \right) \right) - \left( \frac{8}{13} * \log_2 \left( \frac{8}{13} \right) \right) \\
 &= 0,9612366
 \end{aligned}$$

$$\begin{aligned}
 Information\ gain(\text{Total}, \text{Motivasi}) &= 0,90438146 - \left( \left( \frac{12}{25} * 0 \right) + \left( \frac{13}{25} * 0,9612366 \right) \right) \\
 &= 0,4045384
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut penghasilan terlebih dahulu dilakukan seleksi atribut *independent* pada atribut penghasilan, untuk lebih jelasnya dapat dilihat pada tabel 5.

Tabel 5. Seleksi Atribut *Independent* Penghasilan

| Penghasilan | Baru | Bekas |
|-------------|------|-------|
| <1jt        | 5    | 7     |
| 1-5jt       | 9    | 1     |
| >5jt        | 3    | 0     |

$$\begin{aligned}
 Entropy(\text{Penghasilan}, <1\text{jt}) &= - \left( \frac{5}{12} * \log_2 \left( \frac{5}{12} \right) \right) - \left( \frac{7}{12} * \log_2 \left( \frac{7}{12} \right) \right) \\
 &= 0,97986876
 \end{aligned}$$

$$\begin{aligned}
 Entropy(\text{Penghasilan}, 1-5\text{jt}) &= - \left( \frac{9}{10} * \log_2 \left( \frac{9}{10} \right) \right) - \left( \frac{1}{10} * \log_2 \left( \frac{1}{10} \right) \right) \\
 &= 0,46899559
 \end{aligned}$$

$$Entropy(\text{Penghasilan}, >5\text{jt}) = 0, \text{ karena hanya terdiri dari satu kelas}$$

$$\begin{aligned}
 Information\ gain(\text{Total}, \text{Penghasilan}) &= 0,90438146 - \left( \left( \frac{12}{25} * 0,97986876 \right) + \left( \frac{10}{25} * \right. \right. \\
 &\quad \left. \left. 0,46899559 \right) + \left( \frac{3}{25} * 0 \right) \right) \\
 &= 0,24644622
 \end{aligned}$$

Dari penghitungan iterasi 1 diketahui bahwa atribut harga memiliki nilai *information gain* 0,0005438, atribut motivasi memiliki nilai *information gain* **0,4045384** dan atribut penghasilan memiliki nilai *information gain* 0,24644622, sehingga yang akan menjadi simpul akar (*root node*) adalah atribut motivasi. Untuk lebih memperjelas penghitungan iterasi 1 disajikan pada tabel 6.

Tabel 6. Data Hasil Iterasi 1

| Iterasi |             |         | Jml kasus | Baru | Bekas | Entropi    | Gain             |
|---------|-------------|---------|-----------|------|-------|------------|------------------|
| 1       | Total       |         | 25        | 17   | 8     | 0,90438146 |                  |
|         | Harga       |         |           |      |       |            | 0,0005438        |
|         |             | <1 jt   | 13        | 9    | 4     | 0,89049164 |                  |
|         |             | 5-10 jt | 12        | 8    | 4     | 0,91829583 |                  |
|         | Motivasi    |         |           |      |       |            | <b>0,4045384</b> |
|         |             | Atlet   | 12        | 12   | 0     | 0          |                  |
|         |             | Hobi    | 13        | 5    | 8     | 0,9612366  |                  |
|         | Penghasilan |         |           |      |       |            | 0,24644622       |
|         |             | <1jt    | 12        | 5    | 7     | 0,97986876 |                  |
|         |             | 1-5jt   | 10        | 9    | 1     | 0,46899559 |                  |
|         |             | >5jt    | 3         | 3    | 0     | 0          |                  |

Dari tabel 6 dapat diketahui bahwa yang memiliki nilai *information gain* tertinggi adalah motivasi, sehingga yang akan menjadi simpul akar adalah motivasi, ada dua nilai atribut pada motivasi yaitu motivasi atlet dan motivasi hobi, namun pada atribut motivasi atlet sudah selesai karena sudah menghasilkan simpul daun yaitu beli baru, dengan demikian yang akan dilakukan penghitungan lagi adalah pada atribut motivasi hobi

### 3.1.2 Penghitungan Iterasi 2

Pada iterasi ke dua yang menjadi acuan adalah motivasi sebagai hobi, untuk menghitung entropi pada atribut motivasi sebagai hobi terlebih dahulu dilakukan seleksi atribut *independent* pada atribut motivasi sebagai hobi, untuk lebih jelasnya dapat dilihat pada tabel 7.

Tab 7. Seleksi Atribut *Independent* Motivasi Hobi

| Motivasi | Baru | Bekas |
|----------|------|-------|
| Hobi     | 5    | 8     |

$$\begin{aligned}
 Entropy(\text{Hobi}) &= - \left( \frac{5}{13} * \log_2 \left( \frac{5}{13} \right) \right) - \left( \frac{8}{13} * \log_2 \left( \frac{8}{13} \right) \right) \\
 &= 0,9612366
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut harga berdasarkan motivasi sebagai hobi terlebih dahulu dilakukan seleksi atribut *independent* pada atribut harga berdasarkan pada motivasi hobi, untuk lebih jelasnya dapat dilihat pada tabel 8.

Tabel 8. Seleksi Atribut *Independent* Harga Berdasarkan Motivasi Hobi

| Harga  | Baru | Bekas |
|--------|------|-------|
| <1jt   | 4    | 4     |
| 5-10jt | 1    | 4     |

$$\begin{aligned}
 Entropy(\text{Harga}, <1\text{jt}) &= 1, \text{ karena perbandingan dua kelas rasionya sama}
 \end{aligned}$$

$$\begin{aligned}
 Entropy(\text{Harga}, 5-10\text{jt}) &= - \left( \frac{1}{5} * \log_2 \left( \frac{1}{5} \right) \right) - \left( \frac{4}{5} * \log_2 \left( \frac{4}{5} \right) \right) \\
 &= 0,72192809
 \end{aligned}$$

$$\begin{aligned}
 Information\ gain(\text{Hobi}, \text{Harga}) &= 0,9612366 - \left( \left( \frac{5}{13} * 1 \right) + \left( \frac{8}{13} * 0,72192809 \right) \right) \\
 &= 0,06818734
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut penghasilan berdasarkan motivasi sebagai hobi terlebih dahulu dilakukan seleksi atribut *independent* pada atribut penghasilan berdasarkan pada motivasi hobi, untuk lebih jelasnya dapat dilihat pada tabel 9.

Tabel 9. Seleksi Atribut *Independent* Penghasilan Berdasarkan Motivasi Hobi

| Penghasilan | Baru | Bekas |
|-------------|------|-------|
| <1jt        | 0    | 7     |
| 1-5jt       | 4    | 1     |
| >5jt        | 1    | 0     |

$$\begin{aligned}
 Entropy(\text{Penghasilan}, <1\text{jt}) &= 0, \text{ karena hanya terdiri dari satu kelas}
 \end{aligned}$$

$$\begin{aligned}
 Entropy(\text{Penghasilan}, 1-5\text{jt}) &= - \left( \frac{4}{5} * \log_2 \left( \frac{4}{5} \right) \right) - \left( \frac{1}{5} * \log_2 \left( \frac{1}{5} \right) \right) \\
 &= 0,72192809
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy}(\text{Penghasilan}, >5\text{jt}) &= 0, \text{ karena hanya terdiri dari satu kelas} \\
 \text{Information gain}(\text{Hobi}, \text{Penghasilan}) &= 0,9612366 - \left( \left( \frac{7}{13} * 0 \right) + \left( \frac{5}{13} * 0,72192809 \right) + \right. \\
 &\quad \left. \left( \frac{1}{13} * 0 \right) \right) \\
 &= \mathbf{0,683572}
 \end{aligned}$$

Dari penghitungan iterasi 2 diketahui bahwa atribut harga memiliki nilai *information gain* 0,06818734 dan atribut penghasilan memiliki nilai *information gain* **0,683572**, sehingga yang menjadi simpul percabangan adalah atribut penghasilan. Untuk memperjelas penghitungan pada iterasi 2 disajikan pada tabel 10.

Tabel 10. Data Hasil Iterasi 2

| Iterasi |             |        | Jml Kasus | Baru | Bekas | Entropi    | Gain            |
|---------|-------------|--------|-----------|------|-------|------------|-----------------|
| 2       | Total       |        |           |      |       |            |                 |
|         | Motivasi    | Hobi   | 13        | 5    | 8     | 0,9612366  |                 |
|         | Harga       |        |           |      |       |            | 0,06818734      |
|         |             | <1jt   | 8         | 4    | 4     | 1          |                 |
|         |             | 5-10jt | 5         | 1    | 4     | 0,72192809 |                 |
|         | Penghasilan |        |           |      |       |            | <b>0,683572</b> |
|         |             | <1jt   | 7         | 0    | 7     | 0          |                 |
|         |             | 1-5jt  | 5         | 4    | 1     | 0,72192809 |                 |
|         |             | >5jt   | 1         | 1    | 0     | 0          |                 |

Dari tabel 10 dapat diketahui bahwa yang memiliki nilai *information gain* tertinggi adalah penghasilan, sehingga penghasilan akan menjadi percabangan berikutnya, ada tiga atribut yang ada pada percabangan penghasilan, namun pada atribut penghasilan <1jt sudah selesai dan menjadi simpul daun yaitu beli bekas, sedangkan atribut penghasilan >5jt menjadi simpul daun yaitu beli baru, sehingga yang akan dilakukan penghitungan berikutnya adalah atribut penghasilan 1-5jt.

### 3.1.3 Penghitungan Iterasi 3

Pada iterasi ke tiga yang menjadi acuan adalah penghasilan 1-5jt berdasarkan motivasi sebagai hobi, untuk menghitung entropi pada atribut penghasilan 1-5jt berdasarkan motivasi sebagai hobi terlebih dahulu dilakukan seleksi atribut



*independent* pada atribut penghasilan 1-5jt berdasarkan motivasi sebagai hobi, untuk lebih jelasnya dapat dilihat pada tabel 11.

Tabe 11. Seleksi Atribut *Independent* Penghasilan 1-5jt Berdasarkan Motivasi Sebagai Hobi.

| Penghasilan | Baru | Bekas |
|-------------|------|-------|
| 1-5jt       | 4    | 1     |

$$\begin{aligned}
 Entropy(\text{Penghasilan}, 1-5jt) &= - \left( \frac{4}{5} * \log_2 \left( \frac{4}{5} \right) \right) - \left( \frac{1}{5} * \log_2 \left( \frac{1}{5} \right) \right) \\
 &= 0,72192809
 \end{aligned}$$

Untuk menghitung entropi dan *information gain* pada atribut harga berdasarkan motivasi sebagai hobi dan penghasilan 1-5jt terebih dahulu dilakukan seleksi atribut *independent* pada atribut harga berdasarkan motivasi sebagai hobi dan penghasilan 1-5jt, untuk lebih jelasnya dapat dilihat pada tabel 12.

Tabe 12. Seleksi Atribut Independent Harga Berdasarkan Motivasi Sebagai Hobi dan Penghasilan 1-5jt.

| Harga  | Baru | Bekas |
|--------|------|-------|
| <1jt   | 4    | 0     |
| 5-10jt | 0    | 1     |

$$Entropy(\text{Harga}, <1jt) = 0, \text{ karena hanya terdiri dari satu kelas}$$

$$Entropy(\text{Harga}, 5-10jt) = 0, \text{ karena hanya terdiri dari satu kelas}$$

$$\begin{aligned}
 Information\ gain(1-5jt, \text{Harga}) &= 0,72192809 - \left( \left( \frac{4}{5} * 0 \right) + \left( \frac{1}{5} * 0 \right) \right) \\
 &= 0,7219281
 \end{aligned}$$

Untuk memperjelas penghitungan pada itersi 3 disajikan pada tabel 13.

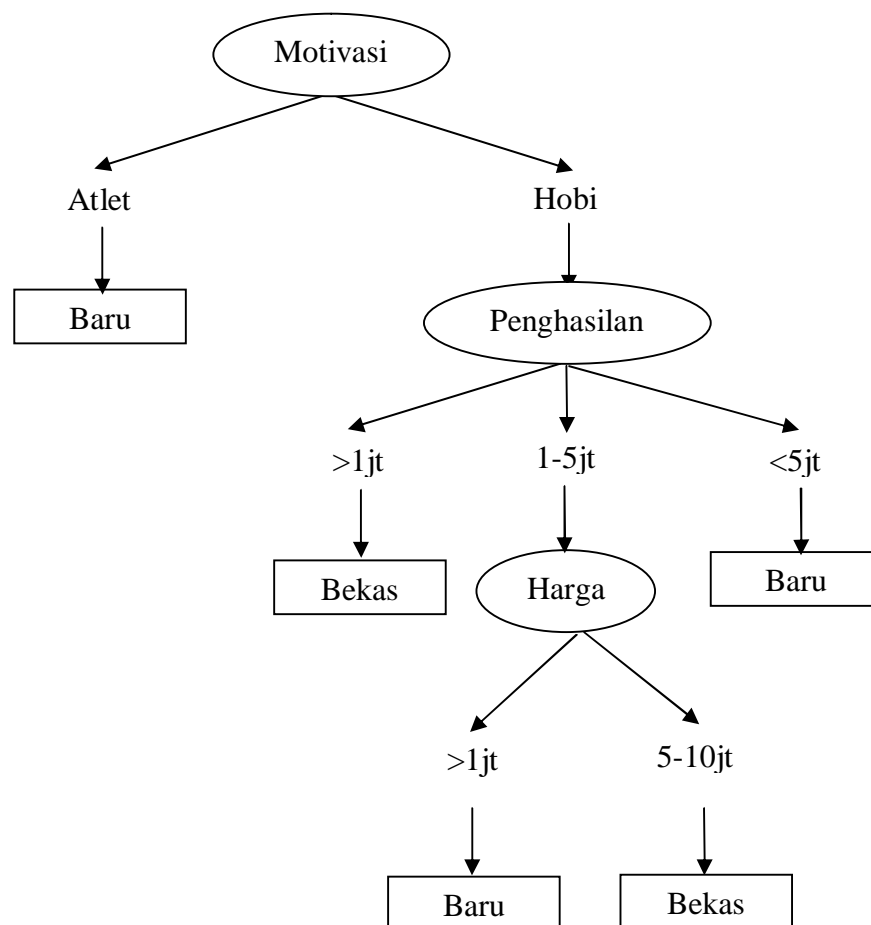
Tabel 13. Data Hasil Iterasi 3

| Iterasi |             |        | Jml Kasus | Baru | Bekas | Entropi    | Gain             |
|---------|-------------|--------|-----------|------|-------|------------|------------------|
| 3       | Total       |        |           |      |       |            |                  |
|         | Penghasilan | 1-5jt  | 5         | 4    | 1     | 0,72192809 |                  |
|         | Harga       |        |           |      |       |            | <b>0,7219281</b> |
|         |             | <1jt   | 4         | 4    | 0     | 0          |                  |
|         |             | 5-10jt | 1         | 0    | 1     | 0          |                  |

Dari tabel 13 dapat diketahui bahwa harga menjadi simpul percabangan terakhir, harga memiliki dua atribut yaitu harga <1jt dengan hasil beli baru dan harga 5-10jt dengan hasil beli bekas.

### 3.2 Bentuk Pohon Keputusan

Setelah dilakukan penghitungan entropi dan *information gain* maka dibuat sebuah pohon keputusan sesuai aturan algoritma C4.5 dalam membangun pohon keputusan. Hasil jadi dari pohon keputusan disajikan pada gambar 1.



Gambar 1. Hasil Jadi Pohon Keputusan

#### 4. PENUTUP

- Data mining klasifikasi algoritma c4.5 dapat membantu penjual alat panah untuk mengetahui perilaku seseorang dalam membeli alat panahan
- Motivasi menjadi faktor utama yang menentukan seseorang untuk membeli alat panahan, diikuti penghasilan dan harga.
- Jika motivasi atlet, maka beli baru.
- Jika motivasi hobi, penghasilan <1jt, maka beli bekas.
- Jika motivasi hobi, penghasilan >5jt, maka beli baru.
- Jika motivasi hobi, penghasilan 1-5jt, harga <1jt, maka beli baru.
- Jika motivasi hobi, penghasilan 1-5jt, harga 5-10jt, maka beli bekas.

#### DAFTAR PUSTAKA

Hermawati, F.A. (2009). *Data mining*. Yogyakarta: Andi.

Kamagi, D.H., & Hansun, S. (2014). Implementasi *data mining* dengan algoritma c4.5 untuk memprediksi tingkat kelulusan mahasiswa. *Ultimatic*, 4(1), 15-20.

Kurniawan, Y.I., & Windiasani, P.A. (2017). Sistem pendukung keputusan untuk penentuan kelulusan beasiswa Sekolah Menengah Kejuruan (SMK) menggunakan metode fuzzy. *Jurnal Teknik Elektro*, 9(1), 13-17.

Prasetyo, E. (2012). *Data mining: konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.

Sunjana. (2010). Aplikasi mining data mahasiswa dengan metode klasifikasi decision tree. *Seminar Nasional Aplikasi Teknologi Informasi*, A-24-A-28.

Supriyono, H., & Sari, C.P. (2015). Pemilihan rumah tinggal menggunakan metode weighted product. *Jurnal Ilmu Komputer dan Informatika*, 1(1), 23-28.

Swastina, L. (2013). Penerapan algoritma c4.5 untuk penentuan jurusan mahasiswa. *Jurnal Gema Aktuaita*, 2(1), 93-97.